

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名: 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密口，在 年解密后适用本授权书

2、不保密□。

(请在以上相应方框内打“√”)

作者签名: 年 月 日

导师签名： 年 月 日

摘 要

近年来,随着我国 P2P 网络借贷的发展,越来越多的学者开始关注借贷关系中存在的歧视现象。本文利用人人贷平台的 128,532 条借贷交易数据,实证研究了 P2P 网络借贷中的地域歧视现象。

研究表明:在控制了借款人的个人信息和借贷合约信息后,各省份之间的借款成功率存在显著的差异,我国 P2P 网络借贷市场上存在着地域歧视现象。进一步研究表明:这种地域歧视现象在借款人的学历和年龄上存在异质性:一方面,借款人的学历越高,越不容易受到地域歧视的影响;另一方面,借款人的年龄越大,P2P 市场上地域歧视的现象越明显。最后,本文研究发现,P2P 市场上出现的地域歧视是非理性的,借款成功率高的省份并没有表现出更低的借款违约率。

关键词: P2P 网络借贷、地域歧视、借款成功率、违约率

Abstract

In recent years, with the development of China's P2P network lending, more and more scholars have begun to study the phenomenon of discrimination in the lending relationship. This paper empirically studies the phenomenon of regional discrimination in P2P network lending by using 128,532 lending data of Renrendai.com.

The research results show that there exists a significant difference in the success rate of borrowing between provinces after controlling the borrower's personal information and loan contract information, so there exists regional discrimination in the P2P network lending market in China. Further research shows that there exists heterogeneity in regional discrimination. Borrowers of low education level and older age suffer more significant regional discrimination than borrowers of high education level and younger age. Finally, the study found that the regional discrimination in the P2P market is irrational, and the provinces with high borrowing success rate do not have lower loan default rates.

Key Words: P2P Lending; Regional Discrimination; Success rate of obtaining loan; Default Rate

目 录

摘 要	I
ABSTRACT	II
1 导论	1
1.1 选题背景、研究目的和意义	1
1.2 文献综述	2
1.3 论文的研究内容、方法及创新点	4
2 实证思路、数据说明和模型构建	6
2.1 实证思路	6
2.2 数据来源和说明	6
2.3 模型构建	10
3 我国 P2P 市场上地域歧视的实证分析	11
3.1 检验地域歧视是否存在	11
3.2 研究地域歧视的异质性	12
3.3 判断地域歧视是否理性	14
3.4 实证模型的稳健性检验	15
4 研究结论和展望	19
4.1 研究结论和建议	19
4.2 不足之处与后续展望	19
致 谢	20
参考文献	21
附 录	24

1 导论

1.1 选题背景、研究目的和意义

随着互联网技术和移动通讯技术的不断普及和发展,互联网金融在市场中的份额不断扩大。Peer-to-Peer(下文简称为 P2P)网络借贷作为互联网金融三大金融模式之一,一方面因其手续简单、门槛较低且不需要担保抵押,为众多的中低收入人群和中小企业提供了灵活的融资方式;另一方面因其投资方式灵活,收益率高的特点吸引了大量的投资者。网贷之家发布的数据显示:2018 年我国网贷累计成交额突破八万亿,与 2015 年相比增加了近六倍。¹

由于通过互联网签订合同的借贷双方互不认识,而借款者又具有信息优势,因此信用风险的控制是 P2P 平台上最重要的问题。根据信息不对称的理论研究成果,可以通过市场的信号显示来缓解信息不对称的问题,这在 P2P 平台中体现在加强借款者信息的披露。事实上,P2P 网贷平台为了消除借贷关系中存在的信息不对称,会在每笔借款订单中公布借款人各方面的信息来降低贷款人的风险,主要包括:(1)借贷合约信息:如借款金额、借款期限和借款利率;(2)个人经济信息:如工资薪酬、是否有房产车产;(3)个人信用信息:如违约次数,信用等级等;(4)个人非经济信息:如教育水平、家庭背景、所在地域等。由于非经济信息容易受到贷款人个人偏好的影响,因此出现了借贷关系中的歧视问题。

所谓歧视是指不平等的看待,在实际生活中每个人都或多或少地存在歧视他人的行为。事实上经济学中对于歧视的问题早有深入的研究,例如劳动经济学家研究过个人非经济信息(例如:肤色、种族、性别、宗教背景、家庭住址、风俗习惯等)对于劳动者在求职、工资、福利等方面是否遭到了不同程度的歧视的问题(姚先国 2004,孙劲悦 2004,章元 2011)。和经济学领域研究歧视的问题相比,在很长一段时间内,在金融学领域内都没有相应的研究成果,其问题在于要收集到可靠的个人微观金融数据十分困难,受访者往往不愿意透露真实的个人信息。但是,随着 P2P 网络平台的兴起,数据收集的问题得以解决(注:P2P 平台为了保证贷款人的利益,会对借款人提供的信息进行逐一审查,因此从 P2P 平台上获取的数据较为可靠),越来越多的经济学家开始关注借贷市场中的歧视问题。

¹ 数据来源: <https://www.wdzj.com/news/yc/3657162.html>

由于我国国土面积较大且人口众多,各地区发展水平、文化习俗、历史背景都有很大的差别,因此地域歧视的问题相对而言较为普遍(黄国萍 2006)。如果在借贷市场中出现地域歧视问题,那么将会降低资金的运作效率,也将造成社会的不和谐。

本文基于人人贷平台 2011-2014 年共计 12 万多条借贷数据,期望达到以下三点目的:(1)实证探讨我国 P2P 网络贷款中是否存在显著的地域歧视问题?

(2)若存在地域歧视问题,那么这种歧视是否具有异质性,即对于不同特征的借款人的影响程度是否有所区别?(3)若存在地域歧视,那么这种歧视是理性的还是非理性的?

由于国内学者研究网络借贷市场中的地域歧视问题较少,本文在前辈研究的基础上,对研究的内容进行了一定的创新,补充了该领域的相关研究,因此具有一定的理论意义;此外,对于借贷双方而言,本文的研究成果能给贷款人理性、准确的选择借款人提供帮助,也能帮助借款人更好的评估自己提供的信息能否获得贷款;对于 P2P 网贷平台而言,本文的研究可以给平台更好的控制违约风险提供理论依据和数据支持,也能帮助平台进行更合理的借贷管理,因此具有实践意义。

1.2 文献综述

在 P2P 网络贷款中,借贷双方互不相识,贷款人很难掌握借款人的真实信息,这使得借款人拥有更大的信息优势,进而引起借贷市场中的逆向选择和道德风险问题(Freedman et al.2011, 谈超等 2014)。为了降低自己的投资风险,贷款人不仅根据借款人的信用等级和个人经济信息来做出投资决策,还常常参考借款人的非经济信息(Freedman et al.2008)。由于贷款人可能对某些事物具有长期负面的舆论导向,因此在借贷市场中容易出现歧视现象(Krumme et al.2009)。早期学者根据歧视产生的原因是否具有经济解释,将歧视分为有效统计歧视和非有效偏好歧视(Becker 1957, Pheleps 1972, Arrow 1973)。具体到我们要研究的网络借贷这一话题,如果贷款人的歧视是由于借款人本身具有较高的违约率而引起的,那么这种歧视就是有效统计歧视。若歧视完全由贷款人的个人偏好决定的,即背后并不存在合理的经济解释,则这种歧视为非有效偏好歧视。

相比于国外金融业发达的国家,我国 P2P 网络借贷起步较晚。在此之前,国外学者对借贷市场进行了各类歧视现象的研究,如社交关系、外貌、种族、年龄、性别、地域等。Freedman 和 Jin(2008)研究了借款人的社交关系对于借款成功率的影响,研究表明,加入 P2P 平台社群的借款人的成功几率高于未加入的借款人。Ravina(2012)研究了美国借贷市场中的外貌歧视问题,研究发现长相漂亮的女性能够获得比男性更低的贷款利率。Pope 和 Sydnor(2011)利用美国 P2P 平台上的借贷数据,发现网络借贷中存在种族歧视,黑人获取贷款的成功率比白人要低 25%-35%,Herzenstein et al.(2011)也证实了这一结论。Pope 和 Sydnor(2011)还发现在借贷市场中存在着年龄歧视现象,35 岁以下的借款者与 35 岁-60 岁的借款者相比,获取贷款更易成功,这与 Barasinska 和 Schafer(2010)的研究发现一致。Schafer(2009)和 Bellucci(2010)的研究均发现,在信贷市场中,女性更容易受到贷款人的歧视,其在借款成功率上比男性要低,且平均要支付的利率也要显著地高于男性,而 Barasinska(2014)利用德国 P2P 平台的数据研究性别歧视,发现控制其他变量后,借款成功率并不受到性别的显著影响,但女性借款人更加倾向于利用更高的利率来吸引贷款人。此外,Aggarwal et al.(2015)从社会信任的角度探讨了在 P2P 市场上产生性别歧视的原因,从而解释了为什么不同地区得到的结论有所区别。针对地域歧视的现象,国外学者也进行了相应的研究。Giannetti 和 Yafeh (2008)研究了借贷双方共同的特征对于借贷成功率的影响,研究发现具有相同性别、种族和地域的双方更容易签订借贷合同,Burtch et al.(2014)利用 2005 年至 2010 年期间发生的 300 多万笔个人贷款交易的数据,发现借款者所在地区收入水平越高,越容易获得借款。

近年来,随着我国 P2P 平台的迅速兴起,越来越多的国内学者开始关注我国 P2P 网络借贷中的歧视现象。李悦雷等(2013)使用拍拍贷 2007 年到 2012 年共计 76259 条数据,发现了在校学生身份认证对借贷成功率具有显著的负向影响。陈霄和叶德珠(2016)使用人人贷的数据研究了 P2P 借贷市场中的性别歧视现象,实证结果表明,性别对贷款的成功率的影响并不显著,但单身女性受到的歧视获得贷款的概率却显著地低于男性,孙武军和刘帆(2017)使用时间更长的数据进行实证研究结果也验证了这一结论。廖理等(2015)考察了 P2P 借贷市场中的学历歧视问题,研究表明借款人学历越高,其违约的概率会更低,但并不会因此提升借款成功率。孙武军和樊小莹(2017)用借款人的工作时间作为从业经历的量

化指标,研究了教育背景、从业经历和贷款成功率三者的关系,实证结果显示:前两者能显著的提高借款者的贷款成功率。庄雷和周勤(2015)借助拍拍贷平台的成交数据,研究了 P2P 借贷市场中身份歧视的问题,研究结果表明,在 P2P 网络借贷中,信息披露的总体质量不高,身份歧视的现象显著存在。郭峰(2017)和蒋彧等(2017)均研究了在网络借贷市场中的婚姻歧视现象,研究发现已婚人士获得贷款的几率显著大于未婚人士,同时其贷款的违约率也较低。目前,国内关于网络借贷市场中的地域歧视研究较少,廖理(2015)实证检验了各省份借贷成功率的差异性,发现在 P2P 网络借贷中存在着偏好性歧视,当他并未对歧视产生的原因进行解释。蒋彧和周安琪(2016)在廖理研究的基础上,将各省份分为三类:高收入地区、中等收入地区和低收入地区,结果显示高收入地区的借款人获得贷款的概率更高。彭红枫(2016)从各省经济发展水平、金融生态环境、教育资源等角度解释了网络借贷中地域歧视存在的原因,研究表明:经济水平高、金融环境好和教育资源多的省份的借款人更容易获得借款。

综上所述,现有文献对 P2P 网络借贷中的各种歧视问题进行了大量的研究,给本文的探究提供了很好的参考。但大部分文献仅局限于探究歧视是否存在,并没有探究歧视中存在的异质性,也没有引入最新论文中发现的控制变量。

1.3 论文的研究内容、方法及创新点

1.3.1 研究内容

第一部分为导论。本部分包括选题背景、研究的意义和目的,并对国内外相关的文献进行了梳理,最后介绍了论文的研究内容、方法以及创新点。

第二部分为实证思路、数据说明和模型构建。本部分首先会给出一个总体的研究思路,后续的实证分析将按照这个思路进行。本部分接下来将介绍数据来源,解释各解释变量的符号和含义,并给出主要变量的描述性统计。最后,本部分将给出后续所有的实证模型的设定。

第三部分是我国 P2P 市场上地域歧视的实证分析。本部分将基于人人贷平台 2011 年-2014 年的数据进行实证研究。首先,本文会运用线性回归模型探究 P2P 借贷市场中存在的地域歧视问题,接着会探讨歧视的异质性,并检验歧视是否理性。在本部分最后,会通过一系列的操作来验证模型的稳健性。

第四部分是研究结论和展望。本部分将汇总实证分析部分得到的结论，并利用得到的结论为借贷双方和 P2P 企业提供相关的建议。最后，将给出本文的不足之处和后续研究的展望。

1.3.2 研究方法

(1) 文献综述法

通过收集国内外研究金融借贷市场上歧视的文章，并对每篇文章进行详细的阅读和整理，最终汇总整理出前辈的研究成果和相应的研究方法，为后续论文的展开打下牢固的基础。同时，收集整理他人文献的过程也能发现该领域有待完善之处。

(2) 描述性统计法

将收集到的数据进行描述性统计，使用图表来方便的展示数据，直观的呈现数据的特征。

(3) Probit 回归分析

当因变量为二值变量时，使用 Probit 回归模型更为合适。本文在稳健性检验部分使用了 Probit 方法，验证了本文实证部分结论的稳健性。

1.3.3 创新点

1. 选题的创新：目前国内关于 P2P 借贷市场中地域歧视的研究较少，目前能查到的公开文献仅有三篇。因此本文的研究内容具有一定的创新性，能为 P2P 网贷市场的健康、稳定的发展提供理论依据和数据参考。

2. 研究内容的创新：首先，已发表的文章都是直接设置各省为虚拟变量，本文在它们的基础上，在稳健性部分将各省份合并到各地理区域（如华中地区、西南地区等），研究划区域的地域歧视；其次，以往的研究对于地域歧视的异质性考察的较少，本文将考察地域歧视在借款人的年龄和学历中存在的异质性；最后，以往的文章在控制变量中没有引入借款人的信息可读性，本文引入了相关变量以消除内生性的问题。

2 实证思路、数据说明和模型构建

2.1 实证思路

如前文所述,本文的目的是探究我国 P2P 市场上地域歧视的问题。因此,本文首先将验证地域歧视的存在性。本文利用各省份借款人的借贷数据,来检验各省份借款人的借贷成功率是否存在显著的差异,如果差异显著,则说明存在着地域歧视的现象。

在验证地域歧视存在后,我们将从以下两个角度来进一步探究这一现象:

(1) 从借款人的特征角度来探究。本文将探究地域歧视是否在借款人中存在异质性,根据已有的研究结果,贷款人对于借款人的学历水平、年龄、婚姻等个人特征也存在歧视,因此我们将利用个人特征对数据进行分组,来重新研究地域歧视在不同组中的差异。本文选取的分组指标有两个,分别为借款人的学历和借款人的年龄。

(2) 从 Becker 等人提出的有效统计歧视(理性的歧视)和非有效偏好歧视(非理性的歧视)角度探究。如果歧视是由于借款人本身具有的较高违约率而引起的,则我们可以认为这种歧视是理性的,因为违约率高可以用来解释地域歧视的产生;否则这种歧视是非理性的,会降低金融市场的运行效率。

2.2 数据来源和说明

本文使用的数据来自陈霄(2018)论文中公开的附件²。数据来自“人人贷”平台,选用“人人贷”平台作为研究对象的三点原因可参见其论文正文。原作者利用爬虫爬取了人人贷网站上2011年1月到2014年6月的所有借款订单数据,并对爬取到的原始数据做了如下处理:(1)将具有机构担保或者实地认证的借款人的订单样本予以剔除;(2)将中国台湾、香港和澳门三地的样本予以了剔除(样本数太少,三地合计样本数为41);(3)将借款者的年龄以及借款额进行了缩尾处理以排除极端值的影响,缩尾的比例为上下1%。经过处理后,本文最终得到的样本数为128532,其中借款成功的样本占比为7.03%,获得借款的借款人的违约率为7.96%。

² 下载地址: <http://www.ciejournal.org/Magazine/show/?id=53529>

下面是具体的数据说明:

1. 因变量

借款是否成功 (*SUCCESS*): 虚拟变量, 对于每一个样本, 如果借款人成功获得借款则记为 1, 否则记为 0.

获得借款后是否违约 (*DEFAULT*): 虚拟变量, 对于每一个样本, 如果借款人(已获得贷款)违约了则记为 1, 否则记为 0.

2. 自变量

(1) 借贷合约信息

取对数后的借款金额 (*LNAMOUNT*): 人人贷平台允许借款人借款的金额范围为3000~500000 元之间, 且必须为50的整数倍, 我们将其取对数后作为自变量。

借款利率 (*INTEREST*): 在2012年之前, 人人贷平台未对借款人设置的借款利率做任何限制, 从2012年开始, 人人贷平台规定允许设定的利率的范围为10% - 24%.

借款的期限 (*MONTHS*): 人人贷平台上允许借款人自行选择借款的期限, 共有 6 个选择: 3, 6, 9, 12, 18, 24 月。

(2) 个人经济信息

月收入 (*INCOME*): 虚拟变量, 人人贷平台将借款人的收入划分为了七个区间, 分别为1000 元及以下、1001 -2000 元、2001 -5000 元、5001 -10000 元、10001 -20000 元、20001 -50000 元以及50000 元以上。本文用 1 表示月收入为10000 元以上的借款人, 用 0 表示月收入不超过 10000 元的借款人。

有无房产 (*HOUSE*): 虚拟变量, 对于每一个样本, 如果借款人有房产则记为 1, 否则记为 0.

有无车产 (*CAR*): 虚拟变量, 对于每一个样本, 如果借款人有车产则记为 1, 否则记为 0.

(3) 个人信用信息

信用评级 (*CREDIT*): 人人贷平台会根据借款人提供的信息和历史借贷情况进行信用评级, 共划分了7个等级, 其中 HR 评级最低(占93.3%), AA 评级最高。本文用 0 表示评级为 HR 的借款人, 其他六种评级用 1 表示。

(4) 个人非经济信息

参加工作的时长 (*WORKTIME*)：虚拟变量，根据借款人工作时长的不同，人人贷平台设置了四个类别：1 年及以下、1—3 年、3—5 年和 5 年及以上。本文用 1 表示工作时长在 3 年及以上的借款人，用 0 表示工作 3 年以下的借款人。

是否已婚 (*MARRIED*)：虚拟变量，如果借款人已婚则记为 1，其他情况（包括未婚，丧偶和离异）记为 0。

年龄 (*AGE*)：人人贷平台上公布的借款人的年龄。

学历 (*EDUCATION*)：虚拟变量，人人贷平台设置了四个类别表示借款人的学历：高中及以下学历、大专学历、大学本科学历和硕士及以上学历。本文将本科及以上学历的借款人记为 1，低于本科学历记为 0。

本文将上述变量进行了描述性统计分析，结果如表 2.1 所示：

表 2.1 研究变量的描述性统计结果

变量	平均数	标准差	最小值	最大值	样本量
SUCCESS	0.0703	0.2557	0	1	128,532
DEFAULT	0.0796	0.0746	0	1	9,037
LNAMOUNT	9.8840	1.4817	8.01	13.12	128,532
INTEREST	16.0327	3.9206	3	24.4	128,532
MONTHS	12.2961	7.9512	1	36.0	128,532
INCOME	0.2483	0.4320	0	1	128,532
HOUSE	0.4112	0.4921	0	1	128,532
CAR	0.2358	0.4245	0	1	128,532
CREDIT	0.0669	0.2499	0	1	128,532
WORKTIME	0.3734	0.4837	0	1	128,532
MARRIED	0.4831	0.4997	0	1	128,532
AGE	32.3512	6.6029	23	54	128,532
EDUCATION	0.2066	0.4049	0	1	128,532

此外，表 2.2 给出了各省份样本数、借贷成功率和违约率的数据。从表中可以看出，内蒙古的借款成功率最低，仅有 3.4%，江苏的成功率最高，达到了 11.2%；但是从违约率的角度来看，内蒙古的违约率略低于平均违约率，而江苏的违约率却略高于违约率，这表明各省份的借款成功率与违约率之间并没有明显的关系，从图 2.1 能直观的观测到这一现象。

表 2.2 各省份借贷样本数、成功率和违约率

省份	样本数	成功率	违约率	省份	样本数	成功率	违约率
安徽	3816	6.94%	11.32%	辽宁	3419	5.15%	5.11%
北京	5288	8.60%	1.32%	内蒙古	1880	3.40%	6.25%
福建	7715	6.73%	6.94%	宁夏	624	3.53%	13.64%
甘肃	1132	8.39%	2.11%	青海	250	6.00%	0.00%
广东	18402	5.84%	6.80%	山东	10917	8.58%	14.09%
广西	3553	5.15%	9.29%	山西	2939	4.15%	7.38%
贵州	1982	8.43%	5.99%	陕西	2646	7.45%	8.63%
海南	875	6.06%	3.77%	上海	4305	6.64%	6.99%
河北	4492	6.08%	5.86%	四川	5979	4.53%	8.12%
河南	4914	8.93%	5.47%	天津	1259	5.88%	10.81%
黑龙江	2551	4.94%	7.94%	西藏	216	3.70%	0.00%
湖北	4512	5.43%	8.57%	新疆	1075	4.19%	2.22%
湖南	4602	5.37%	8.91%	云南	2560	7.07%	7.73%
吉林	1973	7.40%	8.22%	浙江	10410	10.59%	9.26%
江苏	8896	11.23%	8.41%	重庆	2276	3.43%	6.41%
江西	3074	5.63%	4.62%				

注：各省份成功率=该省份获得贷款的样本数/该省份样本总数；
各省份违约率=该省份违约的样本数/该省份获得贷款的样本数

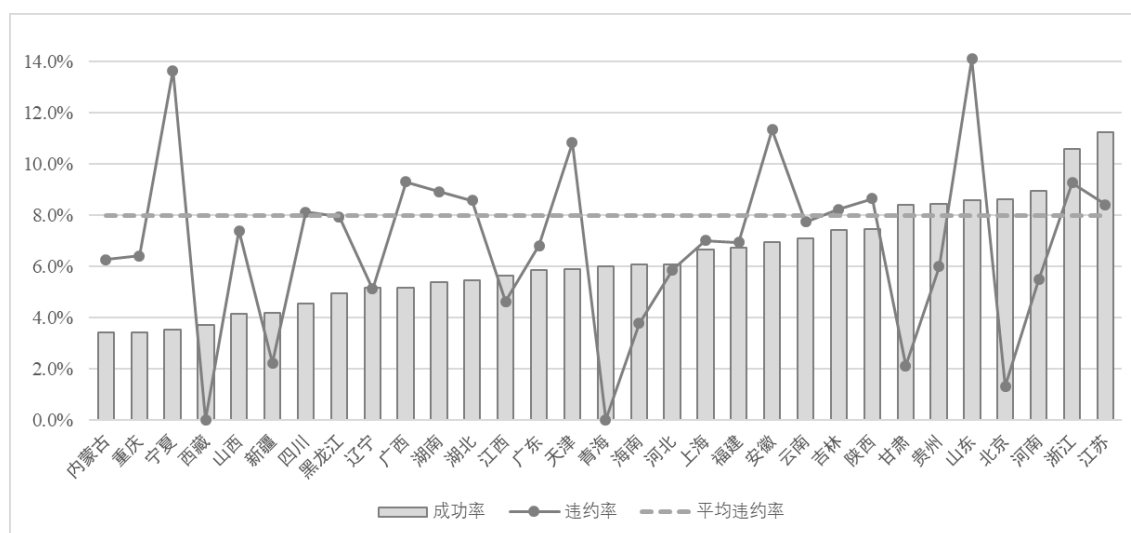


图 2.1 各省份借款成功率、违约率和平均违约率

2.3 模型构建

为了检验 P2P 市场中是否存在地域歧视现象, 本文构建了模型 (1):

$$SUCCESS_i = \alpha + \sum \beta_n \times Province_n + \lambda \times Controls_i + \varepsilon_i \quad (1)$$

这里 $SUCCESS_i$ 表示样本中第 i 个借款人是否获得借款, $Province_n$ 是省份的虚拟变量, 剔除了港澳台三地后, 还剩 31 个省份, 本文设置内蒙古为对照组, 其余 30 个省份设置为虚拟变量, 因此 $n = 1, 2, \dots, 30$, 若第 i 个样本的借款人来自第 k 个省份, 则 $Province_k = 1$, 其余的 $Province_i (i \neq k)$ 全部取 0; 若第 i 个样本来自内蒙古, 则所有的 $Province_n$ 全部取 0。

在加入控制变量后, 本文通过 OLS 估计出所有的系数, 并通过 F 统计量检验回归系数 $\beta_1 = \beta_2 = \dots = \beta_{30}$ 是否为联合显著的, 若联合显著, 则说明存在地域歧视现象。

如果存在地域歧视现象, 我们将进一步探究地域歧视现象, 因此本文建立了模型 (2) - (3)。

模型 (2) 从地域歧视的异质性角度来探究。首先, 利用借款人的某个特征将样本进行分组, 用分组后的数据分别进行回归, 并比较结果的差异。本文采用的分组特征为借款人的学历和工作时间。

假设根据某个特征可将样本分为两个组, 那么模型 (2) 可用下面的式子表示:

$$\begin{aligned} SUCCESS_i^{(1)} &= \alpha^{(1)} + \sum \beta_n^{(1)} \times Province_n^{(1)} + \lambda^{(1)} \times Controls_i^{(1)} + \varepsilon_i^{(1)} \\ SUCCESS_i^{(2)} &= \alpha^{(2)} + \sum \beta_n^{(2)} \times Province_n^{(2)} + \lambda^{(2)} \times Controls_i^{(2)} + \varepsilon_i^{(2)} \end{aligned} \quad (2)$$

模型 (3) 从有效统计歧视和非有效偏好歧视角度出发, 探究 P2P 市场上的地域歧视是否是有效的, 这里使用的样本是全体成功获得借款的借款人, 其样本数为 9037, 因变量 $DEFAULT_i$ 表示第 i 个借款人是否违约。

$$DEFAULT_i = \alpha' + \sum \beta'_n \times province_n + \lambda' \times Controls_i + \varepsilon_i \quad (3)$$

本文通过 OLS 估计出所有的 $\hat{\beta}'_n$ 后, 和模型 (1) 中的估计出来的 $\hat{\beta}_n$ 进行相关性检验, 如果两组系数之间存在显著的负向关系, 则说明地域歧视是理性的, 否则是非理性的。

3 我国 P2P 市场上地域歧视的实证分析

3.1 检验地域歧视是否存在

本文在控制了其它 12 个变量后,将内蒙古设为基准组,使用 OLS 估计模型(1),各省份的回归系数如下表所示(限于篇幅,各控制变量的回归结果参见附件 1):

表 3.1 模型(1)中各省份的回归系数和标准误

省份	回归系数	标准误	省份	回归系数	标准误
上海	0.00471	0.005354	河北	0.00706	0.0050115
云南	0.0228***	0.0060837	河南	0.0249***	0.0052523
北京	0.0123**	0.0052505	浙江	0.0277***	0.0048218
吉林	0.0307***	0.0064521	海南	0.0109	0.0076571
四川	0.00825*	0.0047659	湖北	0.00925*	0.0050054
天津	-0.00312	0.0073158	湖南	0.00624	0.0050286
宁夏	0.0078	0.0082028	甘肃	0.0200**	0.0078838
安徽	0.0235***	0.0054286	福建	0.0114**	0.0047222
山东	0.0170***	0.004691	西藏	0.00844	0.008708
山西	-0.00028	0.0051386	贵州	0.0252***	0.0058424
广东	0.0134***	0.0044345	辽宁	0.00623	0.0051384
广西	0.0122**	0.0052037	重庆	0.000213	0.0054139
新疆	0.00201	0.0068627	陕西	0.0218***	0.0058938
江苏	0.0359***	0.004904	青海	0.0147	0.0129193
江西	0.00259	0.005292	黑龙江	0.0044	0.0054267

Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

从上表可以看出,所有省份的回归系数符号均为正,且有接近一半省份的回归系数在 95%的置信水平上是显著的,这说明这些省份的借款成功率要显著的高于对照组内蒙古的成功率。此外,本文对 30 个省份的系数进行了联合显著性检验,得到的 F 统计量: $F(30,128489)=9.09$,对应的 p 值为 0.00,这说明在 1%的显著性水平上, P2P 借贷市场中存在着地域歧视。

此外,为了直观的展示地域歧视的程度,本文绘制了图 3.1,图上的实线是通过统计性描述计算出来的各省份与内蒙古的成功率的相对差,下方的虚线是模型(1)计算出的回归系数,其实际意义是控制了其他变量(包括借贷合约信息、个人信用、经济与非经济信息)后的相对差。可以直观的发现,尽管差异有所降低,但两条线的整体趋势仍保持一致。从每个省份回归系数的显著性来看,后面差异大的省份更加显著(图中置信水平为 90%),这与直观想法相吻合。

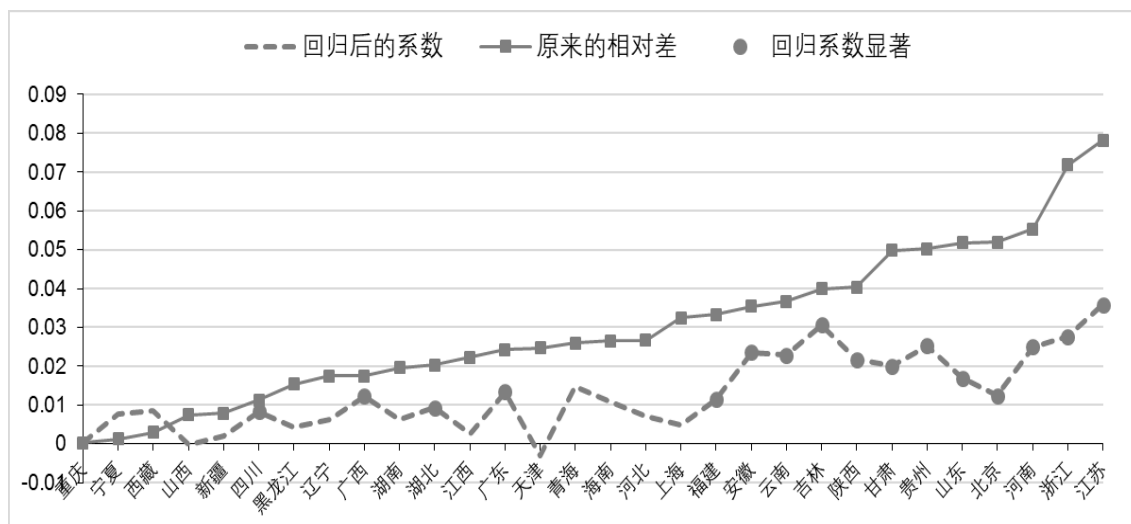


图 3.1 回归后与统计描述各省份成功率相对差

3.2 研究地域歧视的异质性

我们进一步探究地域歧视中存在的异质性，即研究借款人的不同特征对于借款成功率的影响。

首先，我们先利用学历对借款人进行分组，可分为四组：高中及以下学历、大专、大学本科和研究生及以上。利用模型（2）对这四组样本进行回归，限于篇幅，具体的回归结果可见附件 2。从表中大致可以看出，随着学历的提升，回归系数显著的省份的个数会减少。

注意到，学历为研究生及以上的样本数仅有 1832，而前三者中样本数最少的大学本科也有 24,720，因此为了避免较大的误差，我们先对这两个样本进行合并，构造出本科生及以上学历这个分组，这样再重新使用模型（2）进行估计，回归结果可见图 3.2。

图中共有三条折线，分别对应三个分组估计出来的回归系数。此外，为了表示各省份回归系数的显著性（图中所取的置信水平为 90%），图中分别使用了正方形、圆形和三角形标记出了三个分组回归后显著的省份，以大学及以上学历这组为例，仅有云南、吉林、江苏和河南四省的回归系数显著。从图中可以发现，相比于另外两个分组，大学及以上学历这组中，回归系数围绕横轴上下波动，且系数显著的省份的个数远远低于另外两组，这表明借款人的学历越高，越不容易受到地域歧视的影响。

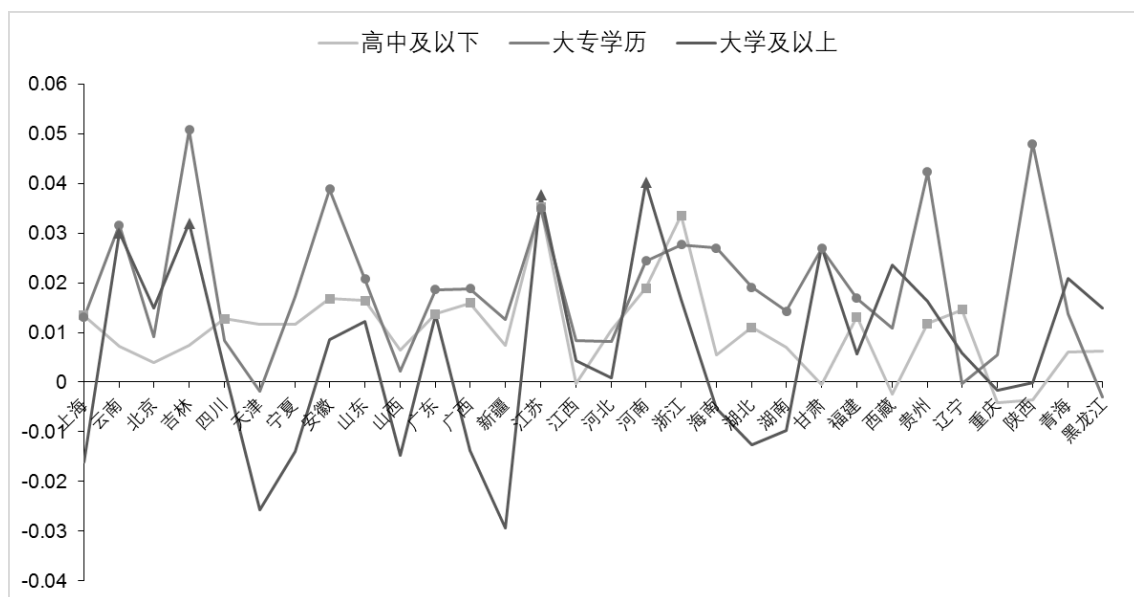


图 3.2 按学历分组的地域差异的回归结果

除了探究地域歧视在借款人的学历中表现的异质性外，本文还探究了借款人的年龄对于借款成功率的影响。图 3.3 是借款人年龄的频数分布直方图，本文根据年龄的大小设置了三个分组，分别为：30 岁及以下、31 岁至 40 岁和 40 岁以上，每个分组对应的样本数分别为 61815、51440 和 15277。

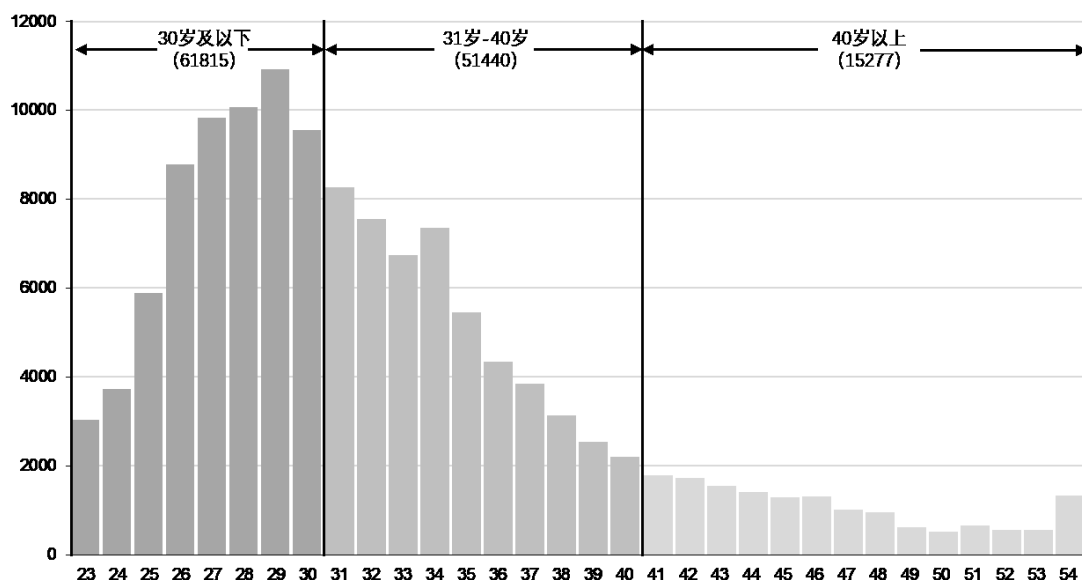


图 3.3 借款人年龄的频数分布直方图

利用模型（2）对这三组样本进行回归，限于篇幅，具体的回归结果可见附件 3。本文我们主要关注不同分组间各省份回归系数的差异和显著性，因此我们绘制了图 3.4，与图 3.2 类似，折线上的标记表示对应的系数显著。从图中明显可以看出，分组为 30 岁及以下的借款人的各省份系数相对于其他两组较小，且并没有出

现显著的省份（在 90%的置信水平上），我们对这一分组中三十个省份的回归系数进行联合显著性检验，得到其统计量 $F(30, 61773) = 1.32$ ，对应的 p 值为 0.1099，这说明在 30 岁及以下的借款人在借款时不存在地域歧视现象。另一方面，从图中可以看出，剩余两个分组均有许多显著的回归系数，对其分别作联合显著性检验可以发现 p 值均接近于 0，且两条折线上均存在较大的省份回归系数，但总的来看 41 岁及以上表现的更加明显。因此我们可以下结论：随着借款人年龄的增加，P2P 市场上地域歧视的现象也更明显。

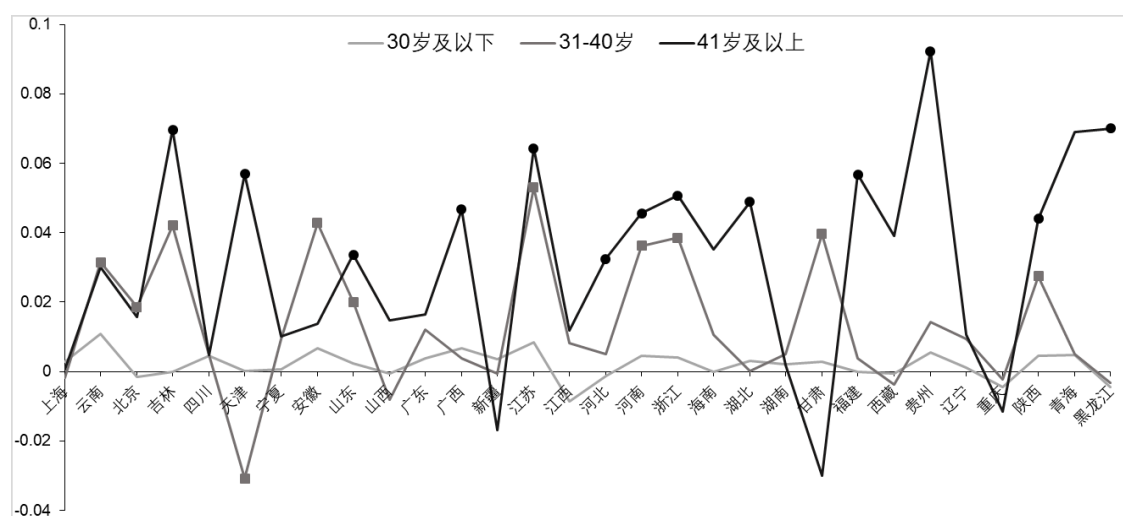


图 3.4 按年龄分组的地域差异的回归结果

3.3 判断地域歧视是否理性

下面我们从另一个角度来探究地域歧视现象，即歧视的分类：有效统计歧视（理性的歧视）和非有效偏好歧视（非理性的歧视）。其核心思想是判断借款成功率低的省份的借款者是否具有更高的违约率。因此，我们可以估计模型（3），其因变量为是否违约，样本为所有成功获得借款的借款人，使用 OLS 可估计出每个省份的回归系数，它们反映了各省份违约率与对照组违约率的相对差额。而我们在模型（1）中也同样估计了各个省份的回归系数，其因变量为借款是否成功，因此其回归系数表示的含义是各省份借款成功率与对照组借款成功率的相对差额。如果这两次回归得到的系数具有显著的负相关关系，则表明地域歧视是理性的；若相关关系不显著甚至具有显著的正相关关系则表明地域歧视是非理性的。

模型（3）的完整回归结果见附件 4，由于我们只关注各省份的回归系数，因此我们在图 3.5 中作出了模型（3）和模型（1）中各省份的回归系数，并用正方形和圆形分别表示两个模型中显著的回归系数。

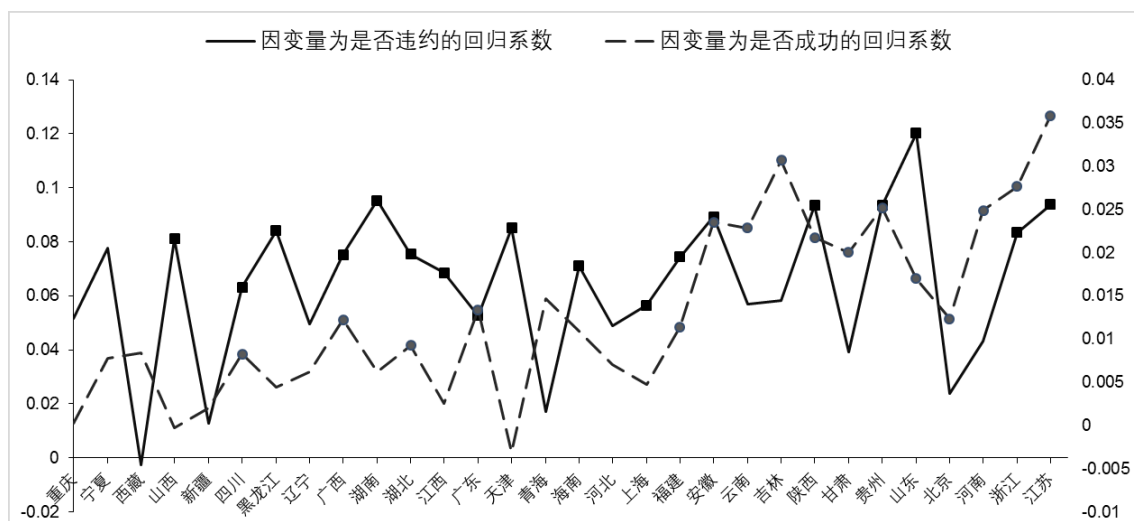


图 3.5 模型（1）和模型（3）各省份系数的回归结果

从上图中我们发现两个模型的回归系数的变化趋势并不稳定，我们可以计算两组数据的样本相关系数 r ,并对其进行费希尔的 t 显著性检验，结果显示两组数据的相关系数为 0.1936，对应的 t 统计量的 p 值为 0.3054，这表明两者没有显著的相关关系。此外为了排除不显著数据的影响，我们将这三十个省份的系数进行了筛选，仅选择两次回归均显著的省份，最终有 11 个省份符合条件，我们再次计算其相关系数为 0.5087，对应的 t 统计量的 p 值为 0.1101，这表明剔除掉不显著数据的影响后，两组回归系数仍然没有显著的相关关系。因此我们可以下结论：P2P 借贷市场中的地域歧视是非理性的，即借款人的违约率高低不是引起贷款人决定投资的主要原因。

3.4 实证模型的稳健性检验

（1）更改地域歧视的研究对象

前文的研究对象是各省份，而各省份的样本数目有较大差异，如广东样本数为 18402，而西藏的样本数仅为 216，两者相差 85 倍，样本数据的不均匀可能会导致回归结果存在一定的误差。限于篇幅，本文在附件 5 中给出了各省份借款成功率的热力图，从图中可以看出，借款成功率在各省份中存在地理上的集聚现象，因此本文按照地理区域将这 31 个省份（剔除港澳台）进行分组。

表 3.2 给出了每个分组中包含的省份，并给出了该组相应的样本数、借款成功率和违约率。

表 3.2 按地域划分后的借贷样本数、成功率和违约率

地域	包含的省份	样本数	成功率	违约率
东北	黑吉辽	7943	5.64%	6.92%
华北	京津冀晋蒙	15858	6.23%	4.35%
华东	苏沪浙皖赣鲁闽	49133	8.71%	9.62%
华南	粤琼桂	22830	5.74%	7.02%
华中	赣鄂湘	14028	6.64%	7.20%
西北	陕甘青宁新	5727	6.53%	6.15%
西南	川渝黔云藏	13013	5.42%	7.23%

从上表可以看出，西南地区省份的借贷成功率最低，而华东地区的借贷成功率最高（注意：其违约率也最高）。

我们将模型一中的省份替换为地域，并选择西南地区作为基准变量，回归结果见附件 6。从回归结果中可以看出，华东地区借款成功率要显著的高于西南地区，而华北地区借款成功率要显著低于西南地区；而其他四个地区的回归系数均不显著。

我们从下面两个角度来说明模型的稳健性：

角度 1：控制变量的回归系数

如下图 3.6 所示，虚拟变量设置为省份或区域对于控制变量的系数大小没有显著的影响，两个模型的回归系数几乎没有差异。

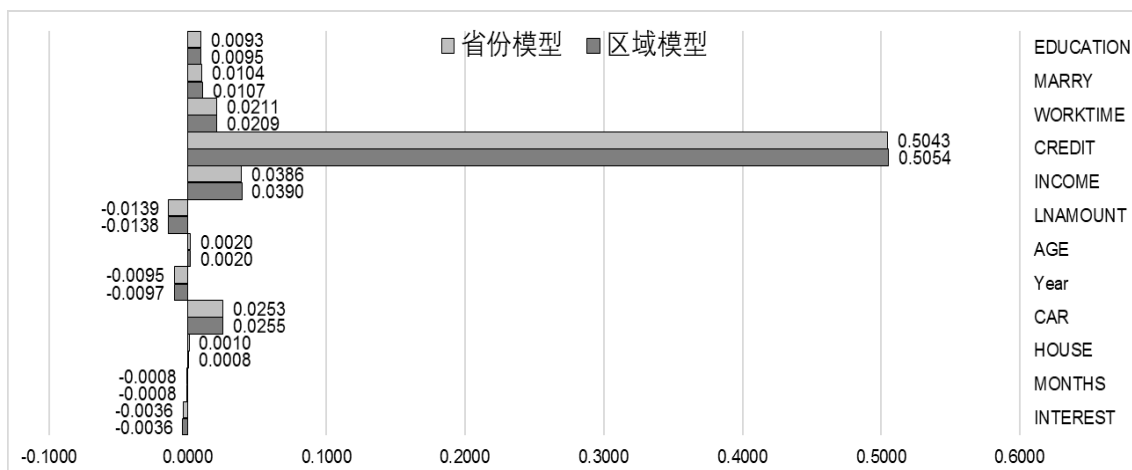


图 3.6 虚拟变量为省份和区域的控制变量的回归系数

角度 2：区域回归系数的联合显著性

我们对模型中六个区域的回归系数进行联合显著性检验，得到的 F 统计量为 11.25，对应的 p 值为 0.00，这说明在 1% 的显著性水平上，P2P 借贷市场中存在着地域歧视，与研究对象为省份时保持了一致。

(2) 更改计量方法

因为本文并不关注借款人所在的省份对于其借款成功率的具体影响大小，所以前文的实证模型采用的均为 OLS 估计。但是，订单的借款成功率为二值变量，为了排除计量方法选择导致的误差，本文重新使用 $Probit$ 模型进行估计。

由于 $Probit$ 是非线性模型，估计量 $\hat{\beta}_{MLE}$ 表示的并不是自变量的边际效应，注意到：
$$\frac{\partial P(y=1|\mathbf{x})}{\partial x_i} = \frac{\partial P(y=1|\mathbf{x})}{\partial (\mathbf{x}'\boldsymbol{\beta})} \frac{\partial (\mathbf{x}'\boldsymbol{\beta})}{\partial x_i} = \frac{\partial \Phi(\mathbf{x}'\boldsymbol{\beta})}{\partial (\mathbf{x}'\boldsymbol{\beta})} \beta_i = \phi(\mathbf{x}'\boldsymbol{\beta}) \beta_i$$
，所以

其边际效用会随着解释变量的变化而变化，为了和 OLS 模型的系数进行对比，我们需要计算平均边际效应，即先计算在每个样本的边际效应，然后再计算其简单算术平均数。

附件 7 中给出了完整的 $Probit$ 回归结果，由于我们关注的是各省份成功率的差异，因此本文作出了 OLS 和 $Probit$ 估计后各省份的回归系数柱状图。从图 3.7 中可以看出，两个模型各省份系数的估计量没有较大差异；另外，图中也展示了各系数的显著性，若系数对应的柱子周围有黑色边框，则代表该系数在 90% 的置信水平上显著，从图中可以看出，与 OLS 回归相比， $Probit$ 回归有三个省份表现的并不显著（四川、湖北和福建），但其他 27 个省份的结果完全一致。附件 7 中同样给出了两个估计方法中控制变量的回归结果对比图，从图中可以看出两个回归模型中控制变量的符号完全一致，除了 $CREDIT$ 这个变量两个模型的估计结果有较大的差异外，其余的 11 个控制变量差异并不明显。同样的，对各省份回归系数进行联合显著性检验，得到的 p 值为 0，这说明了地域歧视现象的存在。

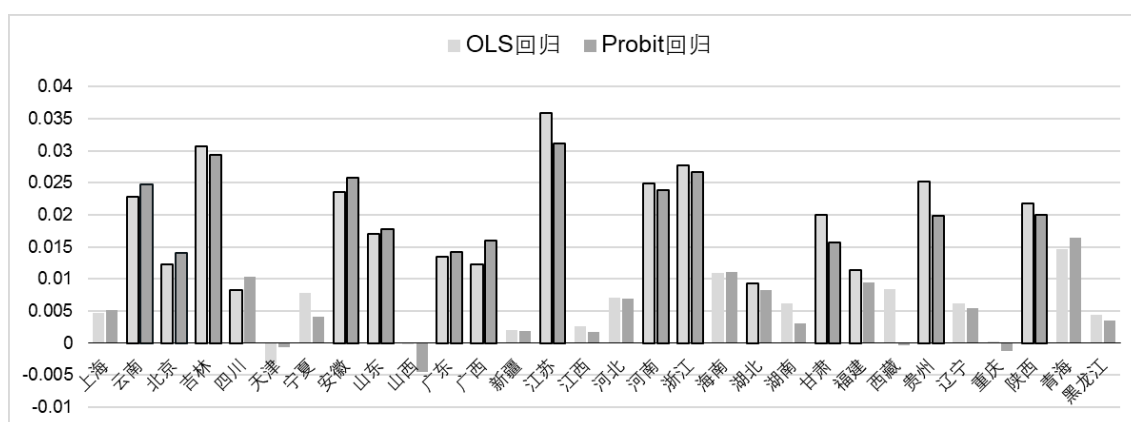


图 3.7 OLS 和 Probit 估计的各省份回归系数

综上所述，本文并没有受到估计方法的影响，得到的结果是稳健的。

(3) 加入其他控制变量

前文的回归模型中引入了 12 个控制变量以消除内生性的影响,但我们忽略了借款人的借款描述信息。事实上,李焰(2014)、陈霄(2018)等人的研究均发现:借款人在借款时描述的可读性能显著提高借款的成功率。因此我们在模型中引入衡量借款人描述可读性的控制变量(使用的控制变量和陈霄(2018)相同),对模型(1)重新进行回归。具体的回归结果见附件 8,与不增加这一控制变量的模型对比,各自变量回归系数的大小和显著性均无明显变化,另外对各省系数进行联合显著性检验,得到 F 统计量: $F(30,128485) = 9.01$, 其对应的 p 值为 0,这说明地域歧视现象仍然存在,因此我们的模型是稳健的。

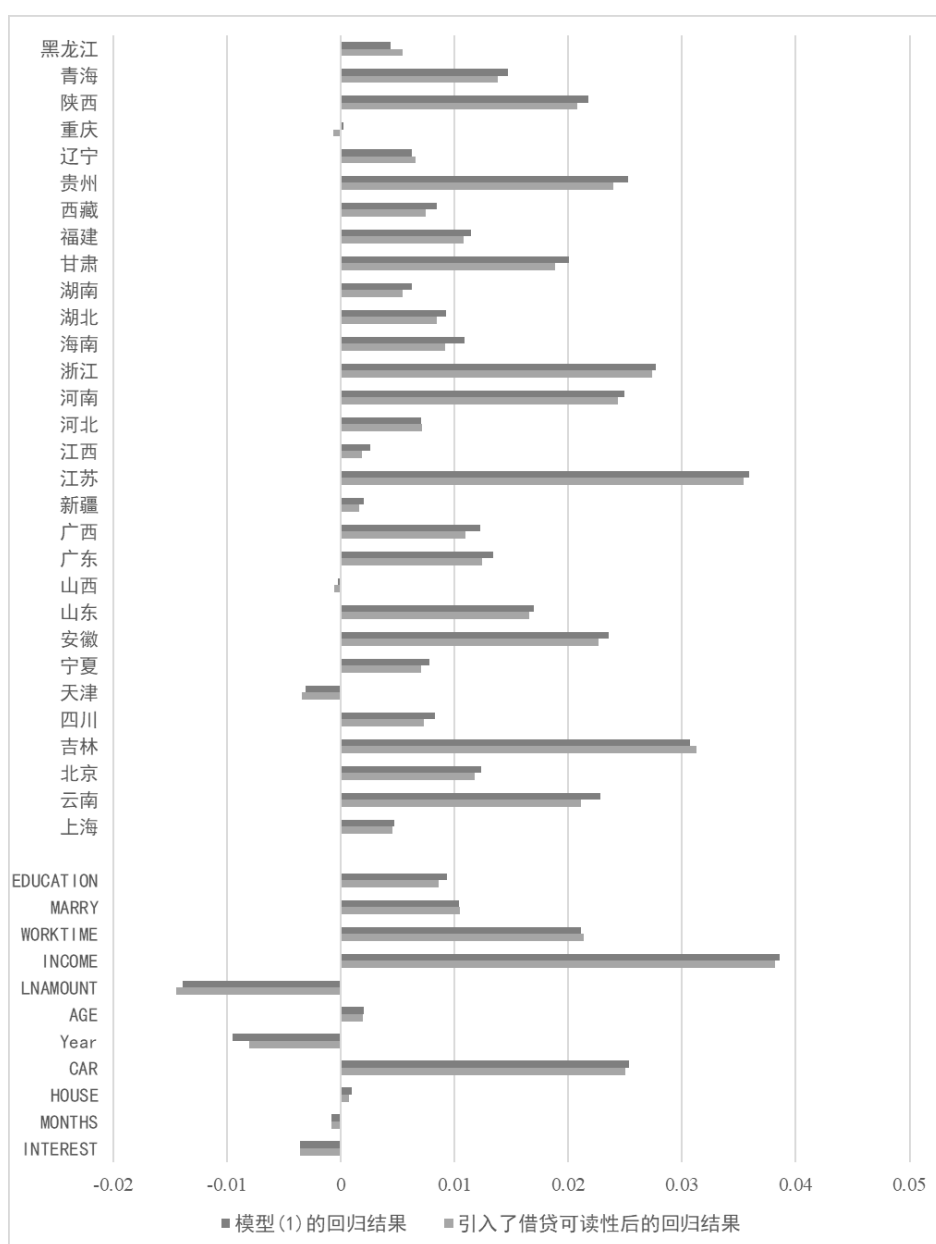


图 3.8 是否加入借贷信息可读性的回归系数对比

4 研究结论和展望

4.1 研究结论和建议

本文利用人人贷平台上借贷交易的数据,研究了我国 P2P 市场上地域歧视的现象,主要得到了以下几点结论:

- (1) 在控制了借款人的信息和借贷信息等变量后,回归结果表明各省份的借款成功率存在显著的差异,且其系数联合显著不为 0,这说明我国 P2P 借贷市场上存在着地域歧视现象;此外,对省份按地域重新分组后进行回归的结果也得到了相同的结论。
- (2) P2P 市场上的地域歧视并不是理性的,借款成功率低的省份的借款者并没有更高的违约率,这说明我国 P2P 平台上存在着“非有效的偏好地域歧视”。
- (3) 我国 P2P 市场上的地域歧视现象对于借款人的学历和年龄存在异质性。一方面,借款人的学历越高,越不容易受到地域歧视的影响;另一方面,随着借款人年龄的增加,P2P 市场上地域歧视的现象也更明显。

本文的建议如下:对于贷款人,在选择借款人时要摒弃地域歧视的观念,更关注借款人的收入、信用等信息,要能根据 P2P 平台提供的各类信息对借款人进行综合判断;对于借款人,要努力提高自身的信用等级,在发布借款信息时要根据自身的情况合理的设计订单信息;对于 P2P 平台,要进一步完善个人信用评级体系,在订单信息的发布中要突出反映影响借款人还款能力的个人信息和反映借款风险的订单信息,从而能够弱化地域歧视的影响。

4.2 不足之处与后续展望

(1) 样本在各省份的分布不够均匀。本文使用的数据来自人人贷官网(2011年1月到2014年6月),共计128532条,由于各省份的数据量存在较大差异,因此回归中可能会出现偏误。

(2) 本文对于地域歧视的原因的探讨还不够深入,在未来的研究中可以引入区域经济发展水平、文化差异,从这些角度去进一步探讨,进而为缓解“地域歧视”现象提供依据。

参考文献

- [1]Aggarwal R , Goodell J W , Selleck L J . Lending to women in microfinance: Role of social trust[J]. International Business Review, 2015, 24(1):55-65.
- [2]Arrow K. The Theory of Discrimination [M], In Discrimination in Labor Markets, ed.Ashenfelter Orley and Albert Rees, Princeton University Press. 1973
- [3]Barasinska N , Sch?Fer D . Is Crowdfunding Different? Evidence on the Relation between Gender and Funding Success from a German Peer-to-Peer Lending Platform[J]. German Economic Review, 2014, 15(4):
- [4]Becker G. The Economics of Discrimination Chicago [M], University of Chicago Press, 1957.
- [5]Bellucci A , Borisov A , Zazzaro A . Does gender matter in bank–firm relationships? Evidence from small business lending[J]. Journal of Banking & Finance, 2010, 34(12):0-2984.
- [6]Bert D’Espallier, Isabelle Guérin, Mersland R . Women and Repayment in Microfinance: A Global Analysis[J]. Social Science Electronic Publishing, 2011, 39(5):758-772.
- [7]Burtch G , Ghose A , Wattal S . Cultural Differences and Geography As Determinants of Online Pro-Social Lending[J]. Mis Quarterly, 2014, 38(3):págs. 773-794.
- [8]Dorothea Schafer,Alexander Muravyev,Oleksandr Talavera.Entrepreneurs' gender and financial constraints: Evidence from international data[J].Journal of comparative economics,2009,37(2):270-286.
- [9]Freedman S , Jin G Z . Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.Com[J]. Social Science Electronic Publishing,2008:8-43.
- [10]Freedman S, Jin G Z. Learning by Doing with Asymmetric Information: evidence from Prosper. com[R]. National Bureau of Economic Research, 2011.
- [11]Giannetti M , Yafeh Y . Do Cultural Differences Between Contracting Parties Matter? Evidence from Syndicated Bank Loans[J]. Social Science Electronic Publishing, 2008.

- [12]Herzenstein M , Sonenshein S , Dholakia U M . Tell Me a Good Story and I May Lend You Money: The Role of Narratives in Peer-to-Peer Lending Decisions[J]. Journal of Marketing Research (JMR), 2011, 48(SPL):S138.
- [13]Krumme K , Herrero-Lopez S . Do Lenders Make Optimal Decisions in a Peer-to-Peer Network?[C]// IEEE/WIC/ACM International Joint Conferences on Web Intelligence & Intelligent Agent Technologies. IEEE, 2009.
- [14]Phelps E S . The Statistical theory of Racism and Sexism[J]. American Economic Review, 1972, 62(4):659-661.
- [15]Pope D G, Sydnor J R. What's in a Picture?: Evidence of Discrimination from Prosper.com[J]. Journal of Human Resources, 2011, 46(1):53-92.
- [16]Ravina E. Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets[J].SSRN Electronic Journal, 2012.
- [17]陈霄, 叶德珠. 中国互联网金融中的性别歧视研究[J]. 金融评论, 2016(2):1-15.
- [18]陈霄, 叶德珠, 邓洁. 借款描述的可读性能够提高网络借款成功率吗[J]. 中国工业经济, 2018.
- [19]郭峰. 婚姻状态与网络借贷中的信用识别——来自“人人贷”平台的经验证据[J]. 当代经济科学, 2017(3).
- [20]黄国萍, 姚本先. 地域歧视与和谐社会的构建[J]. 社会心理学, 2006(4):50-52.
- [21]蒋戡, 施一舟. P2P 网络借贷中的婚姻歧视现象——基于“人人贷”的经验数据[J]. 财经论丛(浙江财经学院学报), 2017(9).
- [22]蒋戡, 周安琪. P2P 网络借贷中存在地域歧视吗? ——来自“人人贷”的经验数据[J]. 中央财经大学学报, 2016(9):29-39.
- [23]李悦雷, 郭阳, 张维. 中国 P2P 小额贷款市场借贷成功率影响因素分析[J]. 金融研究, 2013(7):126-138.
- [24]廖理, 李梦然, 王正位. 中国互联网金融的地域歧视研究[J]. 数量经济技术经济研究, 2014(5):54-70.
- [25]廖理, 吉霖, 张伟强. 借贷市场能准确识别学历的价值吗?——来自 P2P 平台的经验证据[J]. 金融研究, 2015(3):146-159.

- [26]彭红枫,杨柳明,谭小玉.地域差异如何影响P2P平台借贷的行为——基于“人人贷”的经验证据[J].当代经济科学,2016,38(05):21-34.
- [27]孙劲悦.关于就业年龄歧视原因的调查分析[J].财经问题研究,2004(4):81-85.
- [28]孙武军,樊小莹.从业经历和教育背景是否能提高借贷成功率?——来自P2P平台的经验证据[J].中央财经大学学报,2016(3):33-41.
- [29]孙武军,刘帆.网络借贷中存在性别和婚姻歧视吗?——来自人人贷的经验分析[J].产业经济评论(山东),2017:143.
- [30]谈超,王冀宁,孙本芝.P2P网络借贷平台中的逆向选择和道德风险研究[J].金融经济研究,2014(5):100-108.
- [31]王美艳.城市劳动力市场上的就业机会与工资差异——外来劳动力就业与报酬研究[J].中国社会科学,2005(5):36-46.
- [32]姚先国,赖普清.中国劳资关系的城乡户籍差异[J].经济研究,2004(7):82-90.
- [33]章元,王昊.城市劳动力市场上的户籍歧视与地域歧视:基于人口普查数据的研究[J].管理世界,2011(7):42-51.
- [34]曾永明,张利国.户籍歧视、地域歧视与农民工工资减损——来自2015年全国流动人口动态监测调查的新证据[J].中南财经政法大学学报,2018(5):141-150.
- [35]庄雷,周勤.身份歧视:互联网金融创新效率研究——基于P2P网络借贷[J].经济管理,2015(4):136-147.

附 录

附件 1: 模型 (1) 控制变量的回归系数表

被解释变量: SUCCESS			
INTEREST	-0.00355*** (0.000136)	LNAMOUNT	-0.0139*** (0.000544)
MONTHS	-0.000818*** (7.48e-05)	INCOME	0.0386*** (0.00181)
HOUSE	0.000955 (0.00142)	CREDIT	0.504*** (0.00539)
CAR	0.0253*** (0.00185)	WORKTIME	0.0211*** (0.00139)
Year	-0.00952*** (0.000719)	MARRY	0.0104*** (0.00135)
AGE	0.00204*** (0.000114)	EDUCATION	0.00935*** (0.00167)
Constant	19.29*** (1.444)	Observations	128,532
		R-squared	0.334

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

从上表可以看出, 借款人设定的利率、借款金额和借款的期限均对成功率有显著的负向影响; 但借款人的收入、有车有房均对成功率有正向影响。值得注意的是, 这里借款人有房对借款成功率的提升的影响并不显著, 其原因是借款人有房和有车这两个虚拟变量之间存在显著的正相关性 (相关系数为 0.327), 在剔除调 CAR 重新回归后, HOUSE 对借款的成功率存在显著的正向影响 (限于篇幅本文未列出具体的回归结果, 但 HOUSE 的回归系数为 0.0053)。此外, 从借款人的非经济信息来看, 已婚、工作时间长、年龄大和学历高的借款人获得借款的成功率更高, 这与其他学者的研究结果一致。对于借款人的信用评级, 可以发现这一回归系数为 0.504, 这说明贷款人在挑选借款人时非常在意信用评级, 评级越好的借款人获得借款更易成功。

附件 2: 模型 (2) 中以学历进行分组的回归结果

自变量	高中及以下	大专	本科	研究生及以上
INTEREST	-0.00236*** (0.000162)	-0.00393*** (0.000240)	-0.00570*** (0.000390)	0.00159 (0.00179)
MONTHS	-0.000879*** (8.92e-05)	-0.000612*** (0.000130)	-0.000733*** (0.000207)	-0.00328*** (0.000918)
HOUSE	-0.0104*** (0.00194)	0.00612*** (0.00237)	0.00781** (0.00364)	0.0433** (0.0182)
CAR	0.0250*** (0.00264)	0.0252*** (0.00308)	0.0269*** (0.00434)	-0.0566*** (0.0174)
Year	-0.00723*** (0.000897)	-0.00881*** (0.00124)	-0.0144*** (0.00199)	-0.0542*** (0.00871)
AGE	0.00174*** (0.000149)	0.00201*** (0.000198)	0.00247*** (0.000338)	0.00209 (0.00133)
LNAMOUNT	-0.00654*** (0.000721)	-0.0197*** (0.000941)	-0.0181*** (0.00136)	-0.0135*** (0.00464)
INCOME	0.0423*** (0.00266)	0.0420*** (0.00301)	0.0341*** (0.00416)	-0.0504*** (0.0156)
CREDIT	0.519*** (0.0109)	0.479*** (0.00898)	0.510*** (0.00915)	0.423*** (0.0284)
WORKTIME	0.0227*** (0.00185)	0.0228*** (0.00235)	0.00840** (0.00359)	0.0529*** (0.0158)
MARRY	0.00728*** (0.00163)	0.0116*** (0.00235)	0.0122*** (0.00372)	0.0519*** (0.0173)
上海	0.0134* (0.00755)	0.0131* (0.00785)	-0.0204 (0.0141)	0.112 (0.0831)
云南	0.00718 (0.00708)	0.0316*** (0.00950)	0.0184 (0.0168)	0.257** (0.101)
北京	0.00386 (0.00753)	0.00905 (0.00735)	0.000506 (0.0134)	0.167** (0.0802)
吉林	0.00733 (0.00757)	0.0507*** (0.0108)	0.0236 (0.0168)	0.166* (0.0877)
四川	0.0127** (0.00608)	0.00830 (0.00681)	-0.00245 (0.0140)	0.108 (0.0845)
天津	0.0116 (0.0104)	-0.00175 (0.0103)	-0.0267 (0.0200)	0.0642 (0.0868)
宁夏	0.0116 (0.0109)	0.0172 (0.0136)	-0.0295* (0.0167)	0.367* (0.199)
安徽	0.0168** (0.00708)	0.0388*** (0.00803)	0.00783 (0.0157)	0.0285 (0.0870)
山东	0.0164*** (0.00615)	0.0207*** (0.00670)	0.0100 (0.0137)	0.0568 (0.0829)
山西	0.00638 (0.00687)	0.00222 (0.00732)	-0.0184 (0.0148)	0.125 (0.0883)

广东	0.0137** (0.00566)	0.0186*** (0.00639)	0.00978 (0.0134)	0.0995 (0.0805)
广西	0.0159** (0.00666)	0.0188** (0.00799)	-0.0142 (0.0153)	0.00759 (0.0829)
新疆	0.00738 (0.00809)	0.0125 (0.0102)	-0.0324* (0.0190)	0.0697 (0.0790)
江苏	0.0353*** (0.00657)	0.0349*** (0.00710)	0.0344** (0.0137)	0.115 (0.0816)
江西	-0.000198 (0.00670)	0.00829 (0.00811)	0.00313 (0.0149)	0.0249 (0.0779)
河北	0.0105 (0.00638)	0.00809 (0.00759)	-0.00142 (0.0155)	0.0691 (0.0805)
河南	0.0189*** (0.00660)	0.0244*** (0.00773)	0.0225 (0.0156)	0.299*** (0.0877)
浙江	0.0335*** (0.00628)	0.0276*** (0.00708)	0.0145 (0.0141)	0.0885 (0.0857)
海南	0.00539 (0.00972)	0.0270** (0.0128)	-0.00471 (0.0219)	0.00490 (0.0949)
湖北	0.0110* (0.00641)	0.0191** (0.00743)	-0.0140 (0.0142)	0.0423 (0.0818)
湖南	0.00702 (0.00628)	0.0143** (0.00721)	-0.0164 (0.0144)	0.163* (0.0900)
甘肃	-0.000423 (0.0121)	0.0269** (0.0115)	0.0289 (0.0187)	0.0439 (0.104)
福建	0.0132** (0.00601)	0.0169** (0.00715)	0.00161 (0.0143)	0.0700 (0.0835)
西藏	-0.00236 (0.0129)	0.0109 (0.00759)	-0.0169 (0.0191)	0.266*** (0.103)
贵州	0.0118* (0.00691)	0.0422*** (0.00915)	0.0122 (0.0156)	0.0947 (0.0927)
辽宁	0.0146** (0.00688)	-0.000213 (0.00742)	0.00559 (0.0143)	0.0404 (0.0933)
重庆	-0.00417 (0.00650)	0.00555 (0.00817)	-0.00144 (0.0158)	0.0424 (0.106)
陕西	-0.00354 (0.00660)	0.0480*** (0.00920)	-0.00525 (0.0153)	0.124 (0.0940)
青海	0.00604 (0.0183)	0.0138 (0.0217)	0.0223 (0.0263)	-0.0490 (0.0971)
黑龙江	0.00631 (0.00685)	-0.00300 (0.00746)	0.0160 (0.0162)	0.0306 (0.0839)
Constant	14.60*** (1.801)	17.93*** (2.490)	29.19*** (3.994)	109.1*** (17.51)
Observations	54,081	47,899	24,720	1,832
R-squared	0.313	0.299	0.392	0.375

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

附件 3: 模型 (2) 中以年龄进行分组的回归结果

自变量	30岁及以下	31岁至40岁	40岁以上
INTEREST	-0.00187*** (0.000131)	-0.00426*** (0.000261)	-0.00813*** (0.000569)
MONTHS	-0.000296*** (7.88e-05)	-0.00104*** (0.000140)	-0.00201*** (0.000272)
HOUSE	0.00243 (0.00160)	0.00664*** (0.00238)	-0.0176*** (0.00507)
CAR	0.0149*** (0.00230)	0.0257*** (0.00289)	0.0282*** (0.00546)
Year	0.00240*** (0.000757)	-0.0156*** (0.00130)	-0.0377*** (0.00300)
LNAMOUNT	-0.0108*** (0.000560)	-0.0178*** (0.000938)	-0.0102*** (0.00185)
INCOME	0.0265*** (0.00241)	0.0393*** (0.00286)	0.0764*** (0.00528)
CREDIT	0.398*** (0.0121)	0.521*** (0.00721)	0.506*** (0.0112)
WORKTIME	0.0133*** (0.00178)	0.0194*** (0.00220)	0.0512*** (0.00434)
MARRY	0.00762*** (0.00162)	0.0127*** (0.00222)	0.00457 (0.00518)
EDUCATION	0.00947*** (0.00197)	0.0104*** (0.00268)	0.00308 (0.00626)
上海	0.00283 (0.00614)	-0.00174 (0.00952)	0.000757 (0.0169)
云南	0.0108 (0.00722)	0.0315*** (0.0107)	0.0302 (0.0200)
北京	-0.00146 (0.00605)	0.0186** (0.00924)	0.0156 (0.0168)
吉林	-0.000189 (0.00620)	0.0420*** (0.0112)	0.0697*** (0.0226)
四川	0.00456 (0.00567)	0.00478 (0.00856)	0.00481 (0.0148)
天津	0.000279 (0.00720)	-0.0308** (0.0137)	0.0569** (0.0255)
宁夏	0.000748 (0.00923)	0.00940 (0.0144)	0.0100 (0.0292)
安徽	0.00678 (0.00622)	0.0428*** (0.00983)	0.0138 (0.0159)
山东	0.00246 (0.00549)	0.0201** (0.00821)	0.0337** (0.0137)
山西	-0.000647 (0.00608)	-0.00806 (0.00883)	0.0147 (0.0171)

广东	0.00370 (0.00527)	0.0122 (0.00781)	0.0166 (0.0141)
广西	0.00678 (0.00619)	0.00391 (0.00917)	0.0467*** (0.0168)
新疆	0.00367 (0.00780)	-0.000542 (0.0120)	-0.0168 (0.0240)
江苏	0.00834 (0.00570)	0.0530*** (0.00859)	0.0643*** (0.0151)
江西	-0.00853 (0.00593)	0.00829 (0.00943)	0.0119 (0.0173)
河北	-0.00130 (0.00577)	0.00513 (0.00858)	0.0324** (0.0161)
河南	0.00464 (0.00596)	0.0363*** (0.00962)	0.0456*** (0.0154)
浙江	0.00417 (0.00557)	0.0386*** (0.00847)	0.0508*** (0.0147)
海南	-0.000144 (0.00781)	0.0107 (0.0154)	0.0353 (0.0316)
湖北	0.00318 (0.00595)	0.000213 (0.00862)	0.0489*** (0.0161)
湖南	0.00212 (0.00593)	0.00500 (0.00899)	0.00175 (0.0157)
甘肃	0.00276 (0.00900)	0.0396*** (0.0132)	-0.0301 (0.0306)
福建	-0.000180 (0.00543)	0.00385 (0.00839)	0.0567*** (0.0157)
西藏	-0.000480 (0.0108)	-0.00380 (0.0127)	0.0391 (0.0264)
贵州	0.00549 (0.00675)	0.0142 (0.00978)	0.0923*** (0.0202)
辽宁	0.00119 (0.00620)	0.00936 (0.00879)	0.0106 (0.0154)
重庆	-0.00438 (0.00621)	-0.00229 (0.00969)	-0.0116 (0.0238)
陕西	0.00443 (0.00650)	0.0274*** (0.0104)	0.0440* (0.0241)
青海	0.00472 (0.0121)	0.00512 (0.0238)	0.0689 (0.0602)
黑龙江	-0.00434 (0.00635)	-0.00334 (0.00893)	0.0701*** (0.0195)
Constant	-4.696*** (1.520)	31.61*** (2.621)	76.16*** (6.029)
Observations	61,815	51,440	15,277
R-squared	0.185	0.351	0.390

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

附件 4: 模型 (3) 中判断地域歧视是否理性的回归结果

控制变量:

被解释变量: DEFAULT			
INTEREST	0.00317*** (0.000947)	LNAMOUNT	0.0119*** (0.00265)
MONTHS	0.00625*** (0.000618)	INCOME	0.0100 (0.00635)
HOUSE	-0.00854 (0.00645)	CREDIT	-0.173*** (0.00630)
CAR	0.00185 (0.00619)	WORKTIME	-0.0153** (0.00598)
Year	0.0109*** (0.00296)	MARRY	0.00661 (0.00748)
AGE	0.00188*** (0.000499)	EDUCATION	-0.0144*** (0.00534)
Constant	-22.07*** (5.949)	Observations	9037
		R-squared	0.160

各省份的回归系数:

省份	回归系数	标准误	省份	回归系数	标准误
上海	0.0566*	0.03350	河北	0.049	0.03306
云南	0.057	0.03549	河南	0.0433	0.03192
北京	0.0238	0.03101	浙江	0.0835***	0.03131
吉林	0.0584	0.03727	海南	0.0710*	0.04006
四川	0.0632*	0.03397	湖北	0.0754**	0.03450
天津	0.0852*	0.04413	湖南	0.0953***	0.03460
宁夏	0.0777	0.07635	甘肃	0.0392	0.03448
安徽	0.0894**	0.03508	福建	0.0744**	0.03199
山东	0.120***	0.03189	西藏	-0.0027	0.04086
山西	0.0812**	0.03710	贵州	0.0935***	0.03507
广东	0.0528*	0.03116	辽宁	0.0496	0.03407
广西	0.0753**	0.03610	重庆	0.0515	0.04107
新疆	0.0129	0.03784	陕西	0.0936***	0.03531
江苏	0.0938***	0.03139	青海	0.0172	0.04246
江西	0.0686**	0.03421	黑龙江	0.0841**	0.03790

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

附件 5: 各省份借款成功率的热力图和按地域划分的中国地图

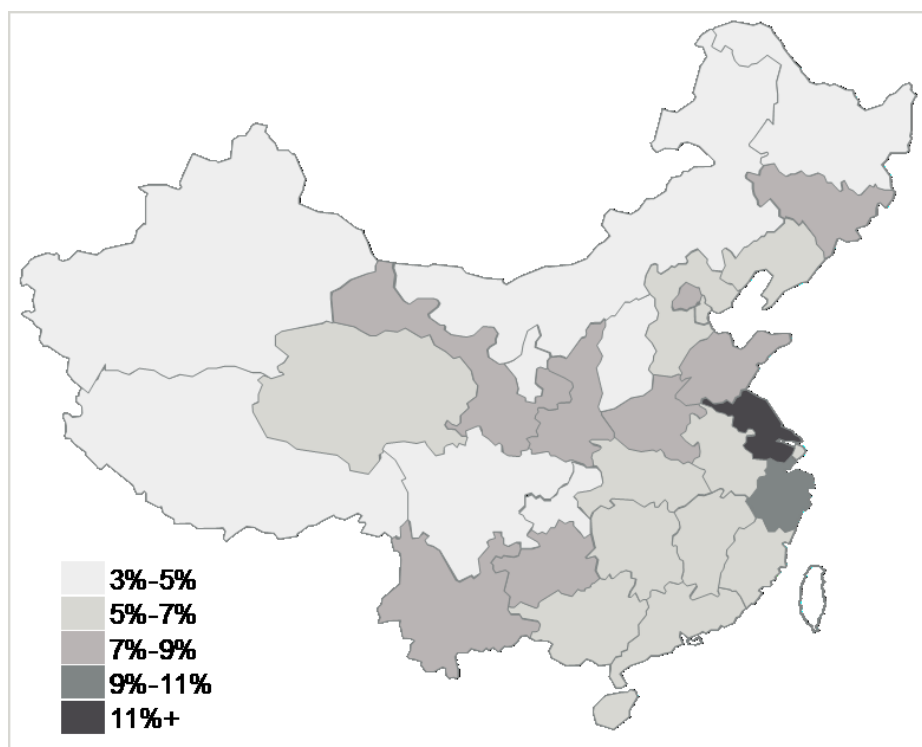


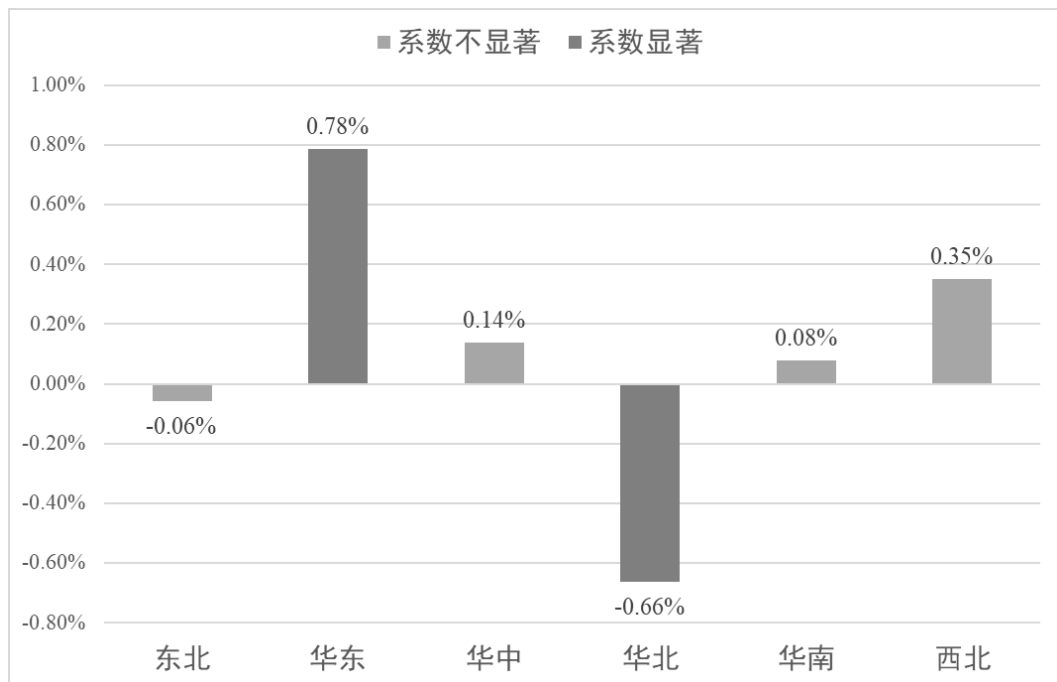
图1: 各省份借款成功率的热力图



图2: 按七大区域划分的中国地图

附件 6: 虚拟变量为区域时模型一的回归结果

被解释变量: SUCCESS			
INTEREST	-0.00360*** (0.000136)	LNAMOUNT	-0.0138*** (0.000543)
MONTHS	-0.000814*** (7.47e-05)	INCOME	0.0390*** (0.00181)
HOUSE	0.000813 (0.00142)	CREDIT	0.505*** (0.00539)
CAR	0.0255*** (0.00185)	WORKTIME	0.0209*** (0.00138)
Year	-0.00968*** (0.000720)	MARRY	0.0107*** (0.00134)
AGE	0.00204*** (0.000114)	EDUCATION	0.00948*** (0.00166)
东北	-0.000592 (0.00267)	华东	0.00785*** (0.00191)
华中	0.00138 (0.00236)	华北	-0.00663*** (0.00228)
华南	0.000788 (0.00209)	西北	0.00352 (0.00318)
Constant	19.61*** (1.446)	Observations	128,532
		R-squared	0.333



附件7: Probit模型回归系数的平均边际效应

控制变量:

被解释变量: SUCCESS			
INTEREST	-0.00307*** (0.0001492)	LNAMOUNT	-0.01233*** (0.0005263)
MONTHS	-0.00122*** (0.0000952)	INCOME	0.03250*** (0.0013721)
HOUSE	0.00188 (0.0013119)	CREDIT	0.12502*** (0.0011561)
CAR	0.01524*** (0.0013385)	WORKTIME	0.01992*** (0.0012624)
Year	-0.00753*** (0.0006044)	MARRY	0.01245*** (0.0014086)
AGE	0.00171*** (0.0000955)	EDUCATION	0.01006*** (0.0013045)

各省份:

省份	回归系数	标准误	省份	回归系数	标准误
上海	0.00519	0.00671	河北	0.00165	0.00715
云南	0.02481***	0.00704	河南	0.00692***	0.00670
北京	0.01411**	0.00646	浙江	0.02384***	0.00647
吉林	0.02934***	0.00741	海南	0.02662	0.00615
四川	0.01032	0.00661	湖北	0.01106	0.00938
天津	-0.00069	0.00846	湖南	0.00823	0.00678
宁夏	0.00409	0.01170	甘肃	0.00301*	0.00682
安徽	0.02577***	0.00669	福建	0.01570	0.00846
山东	0.01780***	0.00617	西藏	0.00950	0.00638
山西	-0.00450	0.00748	贵州	-0.00038***	0.01934
广东	0.01423**	0.00613	辽宁	0.01979	0.00755
广西	0.01597**	0.00696	重庆	0.00542	0.00706
新疆	0.00181	0.00942	陕西	-0.00125***	0.00800
江苏	0.03110***	0.00619	青海	0.01996	0.00709
江西	0.00519	0.00671	黑龙江	0.01643	0.01507

Delta-method standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

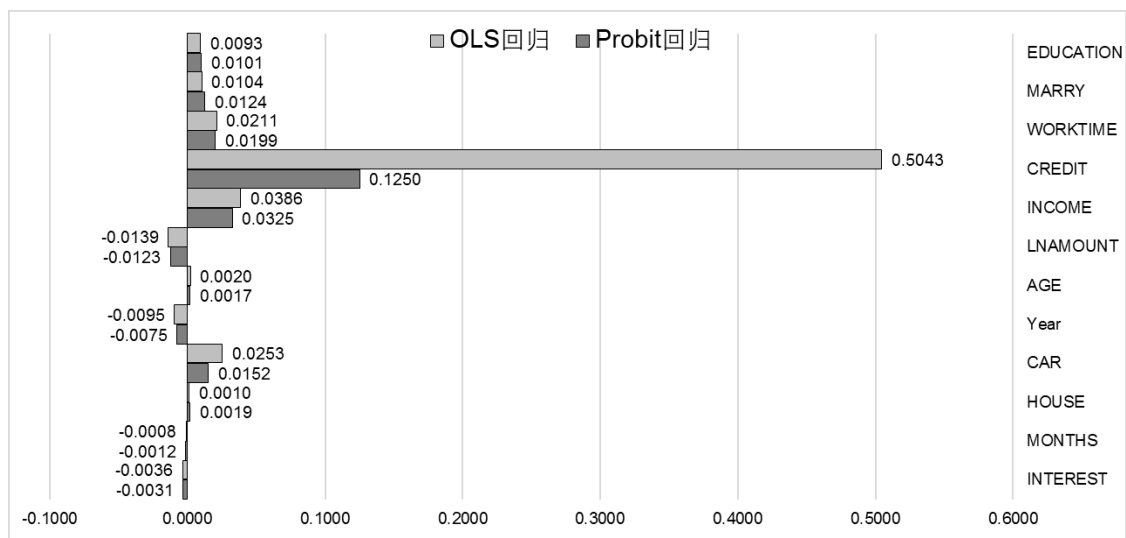


图1: OLS模型和Probit模型的控制变量估计值

附件8: 加入了借贷信息可读性的回归结果

VARIABLES	SUCCESS	SUCCESS
ComWord	0.000196*** (3.36e-05)	
ComVoc	-6.86e-05 (7.78e-05)	
PC_ChiWord	-0.000420*** (0.000105)	
PC_Voc	0.000276 (0.000195)	
ComWord	0.000196*** (3.36e-05)	
MONTHS	-0.000806*** (7.49e-05)	-0.000818*** (7.48e-05)
HOUSE	0.000710 (0.00142)	0.000955 (0.00142)
CAR	0.0250*** (0.00185)	0.0253*** (0.00185)
Year	-0.00803*** (0.000747)	-0.00952*** (0.000719)
AGE	0.00197*** (0.000114)	0.00204*** (0.000114)
LNAMOUNT	-0.0145*** (0.000549)	-0.0139*** (0.000544)
INCOME	0.0382*** (0.00181)	0.0386*** (0.00181)
CREDIT	0.503*** (0.00539)	0.504*** (0.00539)
WORKTIME	0.0213*** (0.00139)	0.0211*** (0.00139)
MARRY	0.0105*** (0.00135)	0.0104*** (0.00135)
EDUCATION	0.00857*** (0.00167)	0.00935*** (0.00167)
上海	0.00455 (0.00535)	0.00471 (0.00535)
云南	0.0211*** (0.00609)	0.0228*** (0.00608)
北京	0.0118** (0.00525)	0.0123** (0.00525)
吉林	0.0313*** (0.00644)	0.0307*** (0.00645)
四川	0.00732 (0.00477)	0.00825* (0.00477)
天津	-0.00342 (0.00732)	-0.00312 (0.00732)
宁夏	0.00703 (0.00821)	0.00780 (0.00820)
安徽	0.0227*** (0.00543)	0.0235*** (0.00543)
山东	0.0166*** (0.00469)	0.0170*** (0.00469)

山西	-0.000561 (0.00514)	-0.000283 (0.00514)
广东	0.0124*** (0.00444)	0.0134*** (0.00443)
广西	0.0109** (0.00520)	0.0122** (0.00520)
新疆	0.00158 (0.00686)	0.00201 (0.00686)
江苏	0.0354*** (0.00491)	0.0359*** (0.00490)
江西	0.00183 (0.00529)	0.00259 (0.00529)
河北	0.00716 (0.00501)	0.00706 (0.00501)
河南	0.0243*** (0.00525)	0.0249*** (0.00525)
浙江	0.0274*** (0.00482)	0.0277*** (0.00482)
海南	0.00917 (0.00766)	0.0109 (0.00766)
湖北	0.00845* (0.00500)	0.00925* (0.00501)
湖南	0.00547 (0.00503)	0.00624 (0.00503)
甘肃	0.0189** (0.00789)	0.0200** (0.00788)
福建	0.0108** (0.00472)	0.0114** (0.00472)
西藏	0.00749 (0.00873)	0.00844 (0.00871)
贵州	0.0239*** (0.00584)	0.0252*** (0.00584)
辽宁	0.00655 (0.00514)	0.00623 (0.00514)
重庆	-0.000665 (0.00541)	0.000213 (0.00541)
陕西	0.0208*** (0.00588)	0.0218*** (0.00589)
青海	0.0138 (0.0129)	0.0147 (0.0129)
黑龙江	0.00546 (0.00543)	0.00440 (0.00543)
Constant	16.29*** (1.500)	19.29*** (1.444)
Observations	128,532	128,532
R-squared	0.335	0.334

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1